DOCUMENT RESUME

ED 306 239 TM 013 002

AUTHOR Kouzekanani, Kamiar; And Others
TITLE Analysis of Testwiseness Components: A Quasi-Experimental Approach.
PUB DATE Mar 89
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
AVAILABLE FROM Kamiar Kouzekanani, P.O. Box 248065, University of Miami, Coral Gables, FL 33124.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *College Bound Students; College Entrance Examinations; Comparative Analysis; Deduction; Guessing (Tests); Higher Education; *High School Graduates; Scores; Test Coaching; Testing Problems; *Test Wiseness; Time Management

ABSTRACT

Differences between four test-independent components of testwiseness and their relative importance were studied in a quasi-experimental investigation of their effects. The components were: (1) time-using training; (2) error-avoidance training; (3) guessing training; and (4) deductive-reasoning training. Three parallel forms of a 30-item test were developed to measure the dependent variable--test performance. High school graduates (n=126) in a 6-week college preparatory program participated in the study. All participants (aged 15-20 years) received training in all components in different order, one each week. A control group received no training. Results indicate that although testwiseness does affect test performance, training in only one component is not sufficient to improve performance. There were no significant differences among the effects of the four components. Members of the training groups started outperforming the control group upon receiving training in at least two components. If only two components can be included in training, one skill should be related to not losing score points and one related to gaining extra points (e.g., error-avoidance and guessing). If three components are chosen, they should be the two skills related to gaining extra points (e.g., guessing and deductive reasoning), and one related to not losing score points (e.g., error-avoidance). A table shows the means and standard deviations of test performance scores. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED306239

Analysis of Testwiseness Components: A Quasi-Experimental Approach

Kamiar Kouzekanani, Maria M. Llabre, and R. Scott Baldwin
University of Miami

Reprints and additional information can be obtained from: Kamiar Kouzekanani, P. O. Box 248065, University of Miami, Coral Gables, FL 33124. (305) 284-3394

Paper Presented at the Annual Meeting of the American Educational Research Associationn, San Francisco, March 1989

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

KAMIAR KOUZEKANANI

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM013002

BEST COPY AVAILABLE

Analysis of Testwiseness Components:
A Quasi-Experimental Approach

Many test-takers receive lower grades on tests than they should because they lack a sophisticated approach to taking tests, or testwisesss (TW). Testwiseness is a multidimensional construct comprised of several component abilities. Testwiseness has been investigated to show that it can be measured (e.g., Gibb, 1964), taught (e.g., Sarnacki, 1979), and that it improves test performance (e.g., Maspons & Llabre, 1985).

The most widely cited definition for TW is the one given by Millman, Bishop, and Ebel (1965). They defined TW as "a subject's capacity to utilize the characteristics and formats of the test and/or the test-taking situation to receive a high score" (p. 707). These authors also provided a taxonomy of TW principles which is comprised of two parts. The first part includes elements which are independent of the test constructor or purpose, namely, time-using, error-avoidance, guessing, and deductive reasoning strategies. The second part includes elements which are dependent upon the test constructor or purpose, they are intent consideration and cue-using strategies. The work of Millman et al. is regarded a seminal work in the area of TW.

The TW components proposed by Millman et al. have been investigated extensively since their inception (e.g., Moore, Schutz, & Baker, 1966; Slakter, 1968; Oakland, 1972; Slaughter, 1975; Goldsmith, 1979; Dreisbach & Keogh, 1982; Bradbard & Green, 1985; Llabre & Froman, 1987). A review of the related literature

revealed that the focus of the investigations has been on those components which are independent of the test purpose or constructor, with several TW techniques manipulated simultaneously. The research strategies exhibited in the literature on TW prohibit the investigation of effects produced by individual components.

The primary purpose of this experimentation was to manipulate the four test-independent components of TW to assess their specific effects on test performance. Four research questions were examined in the study: 1) Are there any significant differences among the four test-independent components of TW? 2) Could training in only one component affect test performance? 3) If training in one component is not enough, how many are needed? 4) What is the most effective order of the four components?

Methodology

This was a quasi-experimental research investigation. There was one independent variable with five levels: time-using training, error-avoidance training, guessing training, deductive reasoning training, and a control group. The training activities were developed based on the works by Heston (1953), Millman and Pauk (1969), and Dobbin (1984).

The dependent variable was test performance. Three parallel forms of a 30-item, 25-minute, objectively scored, and subject-independent test were developed to measure the criterion. The three forms of the test were pilot tested. The coefficients

of equivalence ranged from .83 to .90. A repeated measures analysis of variance revealed no significant differences among the three forms of the test (F (2,36) = .47, p = .63).

Design

The study used a counterbalanced design in which all subjects in the treatment groups received all experimental treatments at some time during the course of the investigation. The experimentation was conducted as a four-week workshop in which one week was devoted to each of the TW components. A one-step cyclic permutation of a sequence of letters was used for the purpose of counterbalancing. The following diagram illustrates the design:

	Week1	Week2	Week3	Week4
T1	A O1	B O2	C O3	D
T2	B O1	C O2	D O3	A
T3	C O1	D O2	A O3	B
T4	D O1	A O2	B O3	C
c	O1	O2	O3	

where: T = Treatment Group, c = Control group, A = Time-Using Training, B = Error-Avoidance Training, C = Guessing Training, D = Deductive Reasoning training, and O = Posttest Measurement. The design of the study made it possible to examine four of the possible 24 permutations of the TW components.

Subjects

One hundred and twenty-six high school graduates, attending a six-week college preparatory program, participated in the study. The subjects ranged in age from 15 to 20 years, with

17.60 as the mean and .70 as the standard deviation. The majority of the subjects were white (62%), followed by Hispanic (25%), and black (13%). Their average SAT score was 886 (sum of verbal and quantitative scores), with 126 as the standard deviation. There were no international students among the subjects, and the majority of them were from middle to upper middle class families. The subjects were assigned to one of the four treatment or control groups based on the availability of time in their schedules, thus, complete randomization was not possible. The five groups were unequal in size due to the time conflict some students had with their other courses. However, there were no significant differences among the five groups with respect to the SAT score, gender, or age of the subjects.

Results

On week one, each treatment group was trained in only one TW component, and the control group received a lecture on educational philosophy and the teaching/learning process. Form A of the subject-independent test was administered to all the participants. A one-way ANOVA revealed no significant differences among the five groups (F (4,121) = .58, p = .67), indicating that training in only one component had no effect on test performance.

On week two, the treatment groups were trained on the second component of TW, and the control group received an orientation regarding the use of the library. All the participants completed Form B of the subject-independent test. A one-way ANOVA showed

significant differences among the five groups (F (4,109) = 7.19, p = .001). Tukey's HSD indicated that the four treatment groups outperformed the control group significantly; however, no significant differences among the treatment groups were observed. T-tests for correlated observations revealed that treatment groups two, three, and four improved significantly from week one; treatment groups one and the control group did not.

On week three, the third component of TW was introduced to the treatment groups. All the participants completed Form C of the subject-independent test. The results of the ANOVA, using Tukey's HSD for the purpose of post hoc analysis, were similar to those reported for week two. The treatment groups did significantly better than the control group, and showed no significant differences among each other; suggesting that the different orders of the components had no significant effect on test performance, (F (4,84) = 8.61, p = .001). T-tests for correlated observations showed that treatment group one improved significantly from week two to week three; the group which did not demonstrate any change from week one to week two. Treatment groups two and three showed some further improvement, but it was not statistically significant. The performance of treatment group four was similar to that observed on week two. The control group remained unchanged.

No test was administered upon the completion of the last week of the workshop, because it became very difficult to develop a valid and reliable fourth form of the test. However, it should

be pointed out that on week four, all the treatment groups had received the training on all four TW components, and it was reasonable to assume that the same results would have been observed (i.e., the treatment groups outperforming the control group and showing no significant differences among each other) if the fourth form of the test had been administered. Nevertheless, the lack of the final form of the measuring instrument was a limitation of the study. There were several absentees during the second and third weeks of the workshop. At no time were significant differences between the absentees and non-absentees observed, based on the week one results, suggesting that the attrition did not bias the results. Table 1 contains a summary of the results

---

Insert Table 1 About Here

Conclusions

At the end of the first week of the study, it was concluded that there were no significant differences among the four TW components; and that training in only one component had no effect on the criterion.

Treatment groups two, three, and four demonstrated significant improvement from week one to week two. Treatment group one, which received training in time-using and error avoidance strategies during the first two weeks of the workshop,

showed no significant improvement. The time-using and error avoidance strategies are designed to assist the test-taker in not losing score points because of reasons unrelated to knowledge of the test content. The greatest improvement belonged to treatment groups two and four which were trained in error-avoidance and guessing, and deductive reasoning and time-using strategies, respectively, during weeks one and two of the workshop. The guessing and deductive reasoning strategies can be used by the test-taker to gain points beyond the sure knowledge of the specific subject matter. Treatment groups two and four were trained in one strategy related to not losing score points and one strategy related to gaining extra score points. Treatment group three, which was trained in guessing and deductive reasoning strategies, showed borderline significant improvement ($p = .04$). The performance of the control group decreased by less than one score point. Based upon the results of the first two weeks of the study, it was concluded that training should include at least two components, and that the most effective combination would be one component related to not losing score points and one component related to gaining extra score points beyond the sure knowledge of the subject matter.

At the end of the third week of the study, the same results were observed. When weeks two and three results were compared, it was found that the significant improvement on test performance belonged to treatment group one only; the group which showed no significant improvement during the second week of the study.

Based upon the analysis of week three data, it was concluded that training should include at least three components if within-group improvement is desired in all the experimental groups. Additionally, since all the treatment groups outperformed the control group, it was concluded that the four different orders of the components examined in this study made no significant contribution to test performance of the participants.

Discussion

The results of this study indicated that although TW does affect test performance, training in only one component is not sufficient, and that there are no significant differences among the four test-independent components of TW. The members of the treatment groups started outperforming the control group upon receiving training in at least two components, suggesting that training should include at least two components. However, within-group improvement was observed in all treatment groups when training was comprised of three components. Ideally, all four components should be included in training, however, if due to some logistical constraints (e.g., time) this is not possible, the following is suggested:

1. If training should include two components, it should be the combination of one skill related to not losing score points and one related to gaining extra score points (e.g., error-avoidance and guessing).
2. If training should include three components, the two skills related to gaining extra score points (i.e., guessing and

deductive reasoning) and one component related to not losing score points (e.g., error-avoidance) are recommended to be included in training.

The participants of this study were a group of high school graduates attending a college preparatory program. One general observation was that high school graduates are not as test-naive as one might expect them to be. Although it has been documented that TW improves test performance, it is not reasonable to assume that training in only one component can accomplish the task. For instance, encouraging test-takers to guess when there is no penalty for guessing without instructing them to utilize deductive reasoning in order to come up with an informed guess may not be fruitful. The treatment groups started to perform significantly better than the control group beginning the second week of the experimentation. That was the time when the new component was related to the one presented the previous week; the two components were synthesized; the students were told about some of the mistakes they had made on the week one test; and the instructor had more to discuss with the class.

Another observation was that just taking tests is not a sufficient means to cause improvement on test performance. In this study, the control group was administered the same tests, and no within-group improvement was observed. Feedback relating the common mistakes to specific TW components is essential if the test-taker is expected to comprehend and apply the strategy to specific testing situations. Sarnacki (1979) advocated that

"mere experience in testing does not guarantee future success on tests, nor does it qualify an examinee as a skilled test-taker" (p. 264).

Regarding the instruction of the TW components, it should be pointed out that with respect to deductive reasoning, in practice, it is inevitable not to mention elements related to intent consideration and cue-using, the two strategies which are dependent upon the test constructor or purpose. There are four reasoning strategies which can assist the test-taker in coming up with an informed guesss. They are absurd options, similar options, opposite options, and give-aways (Sarnacki, 1979). In absurd options, the test-taker is encouraged to eliminate the incorrect alternatives (Gibb, 1964). In similar options, the test-taker should eliminate the two options which convey the same fact because both cannot be correct (Slakter et al., 1970). In opposite options, as suggested by Sarnacki (1979), if there are two options which are opposite in meaning, a sophisticated test-taker can safely eliminate at least one of the options, and can not select both options, since the correctness of one implies that the other one is incorrect. In give-aways, the test-taker could be trained to use the information in other items to select an answer in a present item (Gibb, 1964; Sarnacki, 1979). The deductive test-taker can benefit from these reasoning strategies especially in a poorly constructed test in which cues can be detected by the test-taker.

The time given to complete each test was 25 minutes. On week two, it was observed that the test-takers in the treatment groups took more time to complete the test, and the trend continued during the third week of the study--perhaps because they became more serious about taking the test and put into practice some of the TW skills. For example, it was observed that they were taking advantage of the extra time to review the test, and the number of items skipped by the test-takers was less than the ones observed during the first week of the study.

This study used three parallel forms of a test in which speededness was minimized and the items were sampled from well known standardized tests. Although we did not find any differences among the TW components, it should be pointed out that effects produced by individual components could vary if test items are constrcuted or administered in ways that make them sensitive to those components; for example, speededness or items which are poorly constructed.

REFERENCES

Bradbard, D.A., & Green, S.M. (1985/86). Use of the Coombs elimination procedure in classroom tests. Journal of Experimental Education, 54, 68-72.

Dobbin, J.E. (1984). How to take a test: Doing your best. Princeton, NJ: Educational Testing Service.

Dreisbach, M., & Keogh, B.K. (1982). Testwiseness as a factor in readiness test performance of young Mexican-American children. Journal of Educational Psychology, 74, 224-229.

Gibb, B.G. (1964). Test-wiseness as secondary cue response. Unpublished doctoral dissertation, Stanford University.

Goldsmith, R.P. (1979). The effect of training in test taking skills and test anxiety management on Mexican-American students' aptitude test performance. Unpublished doctoral dissertation, Stanford University.

Heston, J.C. (1953). How to take a test. Chicago: Science Research Associates.

Llabre, M.M., & Froman, T.W. (1987). Allocation of time to test items: A study of ethnic differences. Journal of Experimental Education, 55, 137-140.

Maspons, M.M., & Llabre, M.M. (1985). The influence of training Hispanics in test taking on the psychometric properties of a test. Journal of Research in Mathematics Education, 16, 177-183.

Millman, J., Bishop, C., & Ebel, R. (1965). An analysis of test-wiseness. Educational and Psychological Measurement, 25, 707-726.

Millman, J., & Pauk, W. (1969). How to take tests. NY: McGraw-Hill.

Moore, J.C., Schutz, R.E., & Baker, R.L. (1966). The application of a self-instructional technique to develop a test-taking strategy. American Educational Research Journal, 13, 13-17.

Oakland, T. (1972). The effects of test-wiseness on standardized test performance of preschool disadvantaged children. Journal of School Psychology, 10, 354-360.

Sarnacki, R.E. (1979). An examination of test-wiseness in the cognitive test domain. Review of Educational Research, 49, 252-279.

Slakter, M.J. (1968). The effect of guessing strategy on objective test scores. Journal of Educational Measurement, 5, 217-221.

Slakter, M.J., Koehler, R.A., & Hampton, S.H. (1970). Grade level, sex, and selected aspects of test-wiseness. Journal of Educational Measurement, 7, 119-122.

Slaughter, B.A. (1975). An examination of the effects of teaching and practice in test-taking skills on student performance on a standardized achievement test. Unpublished doctoral dissertation, University of Pittsburg.

Table 1

Means & Standard Deviations of Test Performance Scores

	Week One		Week Two		Week Three	
Group	Mean	SD	Mean	SD	Mean	SD
T1	22.27	3.08	22.14	3.17	24.70	3.04
T2	21.42	3.42	23.58	2.74	25.20	2.31
T3	22.17	3.21	23.84	2.38	24.33	2.22
T4	22.24	3.30	23.47	2.72	23.44	3.16
C	21.26	3.35	19.70	3.23	20.17	3.32

Note: The maximum possible score was 30 points.